

12-2015

American Sign Language Recognition System by Using Surface EMG Signal

Celal Savur
cs1323@rit.edu

Follow this and additional works at: <http://scholarworks.rit.edu/theses>

Recommended Citation

Savur, Celal, "American Sign Language Recognition System by Using Surface EMG Signal" (2015). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the Thesis/Dissertation Collections at RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

American Sign Language Recognition System by Using Surface EMG Signal

by

Celal SAVUR

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of
Science
in Electrical Engineering

Supervised by

Dr. Ferat Sahin

Department of Electrical and Microelectronic Engineering
Kate Gleason College of Engineering
Rochester Institute of Technology
Rochester, New York
December 2015

Approved by:

Dr. Ferat Sahin, Professor

Thesis Advisor, Department of Electrical and Microelectronic Engineering

Dr. Eli Saber, Professor

Committee Member, Department of Electrical and Microelectronic Engineering

Dr. Sildomar T. Monteiro, Assistant Professor

Committee Member, Department of Electrical and Microelectronic Engineering

Dr. Sohail A. Dianat, Professor

Department Head, Department of Electrical and Microelectronic Engineering

Thesis Release Permission Form

Rochester Institute of Technology
Kate Gleason College of Engineering

Title:

American Sign Language Recognition System by Using Surface EMG Signal

I, Celal SAVUR, hereby grant permission to the Wallace Memorial Library to reproduce my thesis in whole or part.

Celal SAVUR

Date

Dedication

I dedicate this work to my loving family who have supported me all the time...

Acknowledgments

Five years ago, I could not have imagined that I would be writing these sentences on my own. Life has such a unique value for everyone and no one knows what it will bring. I would like to thank many who have helped me during my pursuing towards a Master's degree.

I also want to give special thanks to my advisor Dr. Ferat Sahin for always being there and helping me in many of my problems.

Serdar Ciftci and Kamil Bukum, although you are not in the United States, you have helped me a lot. Without your feedback and guidance, I would not be able to accomplish this work.

I am so happy to be a part of the Multi-Agent Biorobotics Laboratory and thank you, everyone, for your help.

Eyup Cinar, thank you so much for your help during my difficult days and guiding me in the right direction.

Shitij Kumar, you are such a good friend and a mentor. I am happy that we will have more time in the future projects during our Ph.D.

Lastly, I would like to thank my brother Thamir EID for always being next to me and supporting me.

For the rest not mentioned **I** thank you **all**.

Abstract

American Sign Language Recognition System by Using Surface EMG Signal

Celal SAVUR

Supervising Professor: Dr. Ferat Sahin

Sign Language Recognition (SLR) system is a novel method that allows hard of hearing people to communicate with society. In this study, an American Sign Language (ASL) recognition system was proposed by using the surface Electromyography (sEMG). The objective of this study is to recognize the American Sign Language alphabet letters and allow users to spell words and sentences. For this purpose, sEMG signals are acquired from subject's right forearm for 27 American Sign Language gestures, 26 English alphabet letters, and one for home position. Time domain, frequency domain (band power), power spectral density (band power), and average power features were used as the feature extraction methods. After feature extraction, Principal Component Analysis (PCA) was applied to obtain uncorrelated features. As a classification method, Support Vector Machine and Ensemble Learning algorithm were used and their performances were compared with tabulated results. In conclusion, the results of this study show that sEMG signal can be used for SLR systems.

List of Contributions

- Implementation of a new input method for users who know American Sign Language.
- Design of a smart phone application that allows hard of hearing people to communicate with hearing people.
- Comparison of results from the Electromyography signal collected for Sign Language Recognition System using four different feature extraction methods: Time domain features, frequency fomain band powers, power spectral density(PSD) band power, and average power in each channel in combination with two different classification metods: Support Vector Machine (SVM) and Ensemble Learning (Bagged Trees).
- **Publication**
Savur, C. and Sahin, F., "Real-Time American Sign Language Recognition Using Surface EMG Signal", in the proceeding of *IEEE International ICMLA Conference*, December 2015, FL, USA.

Contents

Dedication	iii
Acknowledgments	iv
Abstract	v
List of Contributions	vi
1 Introduction	1
2 Literature Survey	3
2.1 Cyber Gloves	3
2.2 Camera and 3D Sensors	5
2.3 EMG Signal	6
3 Background	7
3.1 Preprocessing	7
3.1.1 Zero-Mean	7
3.1.2 Filters	8
3.2 Feature Extraction	9
3.2.1 Time Domain Features	9
3.2.2 Frequency Domain Feature	13
3.2.3 Power Spectral Density	13
3.2.4 Average Power in Channel	14
3.3 Feature Reduction	15
3.4 Classification Methods	18
3.4.1 Support Vector Machine (SVM)	18
3.4.2 Ensemble Learning (Bagged Tree)	21
3.5 Spell Checker	21
4 Proposed Method and Experiments	23
4.1 Data Collection Hardware	23

4.2	Electrode Placements	25
4.3	Psychtoolbox 3.0	26
4.4	Data Collection Process	27
4.5	Offline System Model	29
4.6	Online System Model	31
4.7	Android Application	31
4.8	Spell Checker	32
4.9	Identifiers	33
5	Results and Discussions	36
5.1	Single Subject Results	36
5.2	Multiple Subject Results	42
5.3	Discussions	47
6	Conclusion	49
7	Future Work	51
	Bibliography	53

List of Tables

3.1	Number of the feature before and after PCA	17
5.1	SVM, robustness result with different feature extraction methods	40
5.2	Ensemble Learning robustness result with different feature extraction methods	41
5.3	The sEMG data information of the subjects	42
5.4	Multi users, SVM, robustness result with different feature extraction	43
5.5	Multi users, Ensemble Learning, robustness result with different feature extraction methods	44

List of Figures

2.1	The representation of letter 'u' and 'v' in American Sign Language	5
3.1	Windowing Diagram	8
3.2	Power using absolute value of FFT and PSD using Welch's method	14
3.3	The Human forearm muscle anatomy	14
3.4	A plot of the sEMG of letter 'A' and 'B'	15
3.5	Support Vector Machine	19
3.6	Test result vs. Cross Validation result and best sigma value	20
3.7	Ensemble Learning	21
4.1	The Bio Radio 150 user Unit	24
4.2	The Myo Armband	25
4.3	Bio Radio 150 electrodes placement	26
4.4	Myo Armband device placement	26
4.5	Data Collection Diagram	27
4.6	A custom software that read sEMG signal from device	28
4.7	A screenshot of the simulation that subject performs the letter 'B'	29
4.8	Offline American Sign Language System Model	30
4.9	Online American Sign Language System Model	31
4.10	Smart Phone Application	32
4.11	Spell Checker Diagram	33
4.12	Double tap gesture.	34
4.13	(A) Wave in , (B) Wave out and (C) Fist gestures	34
4.14	How The System Works	35
5.1	The performance of the Ensemble learning using 200 trees	38
5.2	Single subject different feature with SVM classifier	39
5.3	Single subject different features with Ensemble classifier	41
5.4	Different feature extraction methods with SVM classifier with multi subjects	43
5.5	Multi subjects different features with Ensemble classifier	44
5.6	The Confusion matrix of the Single Subject Result	46
5.7	Gesture of letter 'G' and 'H'	47

5.8	Gesture of letter 'R' and 'U'	47
5.9	2D (radomly selected two features) representation of the 27 classes dataset	48

Chapter 1

Introduction

According to the World Health Organization, there are approximately 360 million people having the disability of hearing in 2015 [1]. The number was 278 million in 2005 [2]. In 10 years, the number of people having hearing problems has increased about 14%. Sign Language is the only method hard of the hearing use in their daily life. Sign Language is a hand gesture language that the hearing impaired use to express their thoughts, feeling, and knowledge instead of verbal communication. There are different sign languages available in the world which have their own alphabet and hand gestures. American Sign Language (ASL) is one of the most popular sign language in the world. According to Tamar [3], ASL is the fourth most popular language in the USA and it is trending in college students. The hard of hearing need a communication method which allow them to communicate with others in order to accomplish their daily tasks. Therefore, it is really important to have a Sign Language Recognition (SLR) system.

In this study a novel SLR system is implemented by using sEMG signal. The sEMG signal is the differential voltage in the muscles. The advantage of the sEMG signal is that it is not affected from weak lit areas and easy to use when other methods have limitation such as lighting. The proposed method uses sEMG data to predict American Sign Language

alphabets. With this novel method, the hard of hearing will be able to type any word and sentences to communicate with others.

Chapter 2

Literature Survey

This chapter presents a literature review of the American Sign Language and Machine Learning algorithms used in this study. Sign Language Recognition (SLR) is a topic which has gained a lot of attention by researchers recently. There have been many different approaches proposed to solve this problem. Existing approaches may be divided in to three main groups: Cyber Gloves, Camera and 3D Sensors, and sEMG signal.

2.1 Cyber Gloves

In 1993, Fels et al. were one of the first research group who worked on a Sign Language Recognition system [4]. The authors came up with the idea of a Glove-Talk that acquired data from Cyber Glove, which was later used to recognize text. Neural Networks (NN) were used as the classification technique [4]. In comparison to recent advances, and the lack of technology twenty years ago, the research had impressive results [4]. The SLR system produced wrong words less than 1% of time in 203 gesture vocabularies [4]. Two years later, Liang (1995) et al. used the Cyber glove to recognize ASL alphabet rather than words [5]. The researchers did not use any classification algorithm, instead, they used rule-based voting algorithm. The experiment had satisfactory results, which recognized 3 to 4 characters per second, although the researchers did not use any classifier. Meanwhile,

Jiangqin et al. published the paper about Chinese Sign Language Recognition system [6]. In 2002, Mehdi et al. attempted to develop an ASL system without training their system [7]. The authors ignored letter 'J' and 'Z' because Cyber Glove captures only the shape of the hand and both letters required wrist motion. Subsequently, a Neural Network based classifier was used to discriminate 24 letters and two special characters (26 classes). The researchers did not use any preprocessing (filtering) method, thus acquired signal from gloves were directly fed into the system [7]. As a result, the system had lower accuracy than the trained system. Sole et al. used Extreme Learning Machine (ELM) algorithm to classify the Auslan (Australian Sign Language) alphabet [8]. The reason they chose ELM because it is a simplified version of the Neural Network (NN) [8]. According to reported results, the system had 95% accuracy. Hayek et al. used an efficient way to show the results of the system displayed on LCD [9]. Similar to [5], the author of [9] did not use any classification algorithm. Instead, they used their own binary algorithm and did not share any performance result. The researchers did not mention how the letters that have a similar shape such as letter 'U' and 'V' were handled. The only difference between those letters is the angle between index finger and middle finger as shown in Figure 2.1 [10]. Using their binary approach will not differentiate those letters from each other. In addition, they did not mention how they handled letter 'J' and 'Z'. The methods discussed above are using Cyber Glove for data acquisition. Although, those studies have valuable results, it is not comfortable to wear the Cyber Glove for daily life until there are great advances in wearable technology.

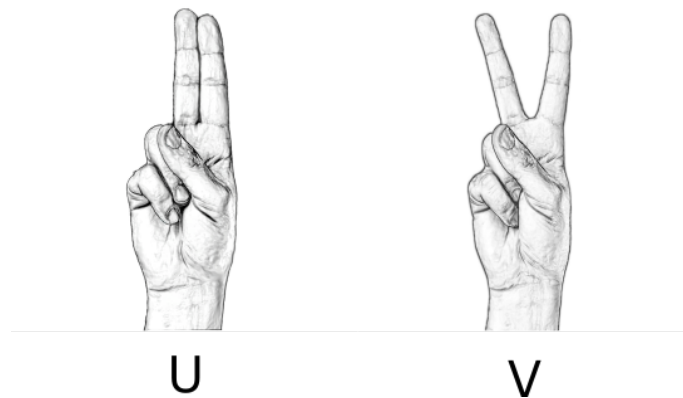


Figure 2.1: The representation of letter 'u' and 'v' in American Sign Language

2.2 Camera and 3D Sensors

Research into Computer Vision has a long history. Therefore, there are many different applications of Computer Vision in SLR systems. In this section, most related studies are mentioned.

In 1998, Starner et al. mounted a camera to a cap in order to capture hand tracking [11]. They used Hidden Markov Models (HMM) to recognize sentence-level ASL rather than explicitly modeling the fingers. The researchers concluded that having a camera mounted on a cap had better accuracy than having desk mounted camera [11]. Otiniano-Rodriguez et al. used the Kinect sensor for ASL recognition system [12]. They mainly focused on comparing RGB, Depth, and RGB-Depth performances on ASL system. They found that combining RGB and Depth information gave better results than single information system [12]. Keskin et al. also used the Kinect sensors [13]. The authors used a Random Decision Forest (RDF) and had satisfactory result of 99.9%. Results of RDF and ANN were compared and tabulated [13]. Chuan et al. developed an SLR system using Leap Motion sensor

[14]. They used Leap Motion sensor because it is easier to transport and economical in comparison to the Kinect sensor and Cyber glove. They used the k-nearest neighborhood and SVM to classify 26 letters of American Sign Language. Meanwhile, Raut et al. used the dataset of 312 ASL hand gestures images [15]. The LGB Vector Quantization method was used to solve SLR system [15]. However, in those experiments, they were using RGB information in SLR system. Therefore, in poor lighting conditions, these methods failed.

2.3 EMG Signal

The third approach is surface Electromyography (sEMG) signals. It is usually used by researchers to control a robot or a prosthetic arm [16] [17]. In addition, sEMG signal was used to increase the usability of Human Computer Interaction (HCI) [18][19]. Erkilinc and Sahin [20] designed a surveillance camera control system for people who cannot use joysticks. In order to extract features, Fast Fourier Transform (FFT) was applied to the raw sEMG data. They also used the Principal Component Analysis (PCA) to eliminate uncorrelated features [20]. Sahin et al. implemented a system, which controls the computer cursor, using sEMG signal [21]. They used wavelet transformation with PCA to extract features and fed them into SVM classifiers. Kosmidou et al. classified most common ten English words without explicitly classifying the letters [22].

Unlike Cyber Glove, camera, and 3D sensor that are restricted to well lit areas, sEMG signals are not affected by light. There is one difficulty for acquiring sEMG signal, i.e. having too many wires around the arm. Thanks to the advance wearable technologies, such as Myo Armband, acquiring sEMG signals has become a lot simpler and less clumsy [23].

Chapter 3

Background

This chapter gives background information about the methods and algorithms used in this study. We will present algorithms and methods for preprocessing, feature extraction, feature reduction, and classification.

3.1 Preprocessing

It is the nature of the raw signal that contains noise, artifact, and mixed with an unwanted signals. For obtaining the desired signal, the raw signal needs to be processed to clear these unwanted elements. This operation is called preprocessing in Machine Learning. In this study, zero mean and band pass filters are applied to the raw data to remove artifacts and unwanted information.

3.1.1 Zero-Mean

Raw signals usually contain unwanted trend or bias. Therefore, analytical methods translate data such that either individual feature have no bias or their mean is zero. The zero meaned

data help to reduce DC component that causes a problem during Fast Fourier Transformation. For calculating zero mean of the raw data, the equation 3.1 was used.

$$X_{zero} = X - mean(X) \quad (3.1)$$

3.1.2 Filters

The first step after obtaining the raw data was to filter the raw sEMG data. Samples of each letter were transformed into a time window which is a common technique for sEMG processing. Figure 3.1 illustrates the windowing technique. The data flow from right to left as shown in figure via arrows. Whenever new data is acquired, it pushes the temp data out of the window. The windows are two seconds long, so each new data frame stays eight cycles in the window. Then, two filters were applied to the zero meaned dataset. The first filter was a band-pass filter consisting of a high-pass filter with 5 Hz cutoff frequency and a low-pass filter with 100 Hz cutoff frequency. The second filter was a notch filter with 60 Hz to remove the power line noise. Remaining data is preprocessed valuable sEMG information which is used in the following feature extraction steps.

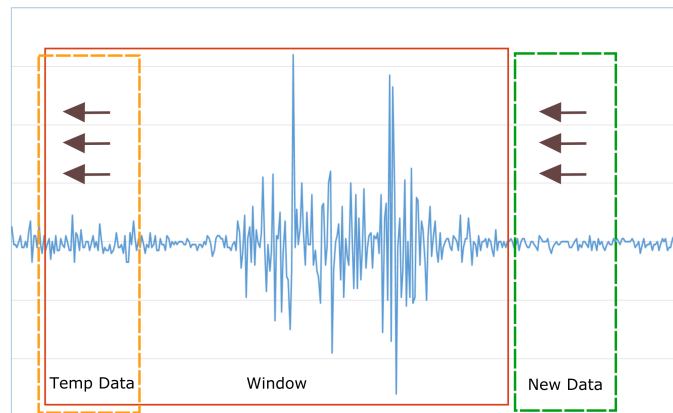


Figure 3.1: Windowing Diagram

3.2 Feature Extraction

In this section, we will discuss some of the feature extraction methods such as: mean absolute value, modified mean absolute value, simple square integral and so on.

3.2.1 Time Domain Features

In machine learning, having distinguishable features are the most important keys as they improve the accuracy of the system significantly. In order to have a good SLR system, following feature extraction techniques were used in this study:

Mean Absolute Value

The Mean Absolute Value (MAV) is a way to represent the signal in the time domain that makes easy to see the pattern of the muscle activity, given in equation 3.2. Since each letter in American Sign Language has different gesture, the MAV might increase the accuracy of SLR system.

$$MAV = \frac{1}{N} \sum_{i=1}^N |X_i| \quad (3.2)$$

In the equation 3.2, X represents a channel information, X_i is a data point in the channel, and N is the number of the samples. In our experiment N is 400.

Modified Mean Absolute Value

The Modified Mean Absolute Value (MMAV) is the extension of the MAV where it uses weighted windows. With this feature, we can set different weights for a certain range of signal amplitude. The equation 3.3 was used, where W_i is a range depends on i .

$$MMAV = \sum_{i=1}^N W_i |X_i|; W_i = \begin{cases} 1, & \text{if } 0.25N \leq i \leq 0.75N \\ 0.5, & \text{otherwise} \end{cases} \quad (3.3)$$

Simple Square Integral

The Simple Square Integral (SSI) is a representation of energy in the sEMG signal [24], given in equation 3.4.

$$SSI = \sum_{i=1}^N |X_i|^2 \quad (3.4)$$

Root Mean Square

Similar to the MAV, the Root Mean Square (RMS) reflects the activity of muscles. According to Phinyomark et al. the features in the RMS are more advanced than the MAV [24]. Therefore, both methods were used in the feature extraction step. This feature calculated by using equation 3.5.

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N X_i^2} \quad (3.5)$$

Log Detector

Log detector provides an estimate of exerted muscle force [25]. Equation 3.6 used for calculation.

$$LOG = e^{\frac{1}{N} \sum_{i=1}^N \log(|X_i|)} \quad (3.6)$$

Average Amplitude Change

The Average Amplitude Change (AAC) measures an average of the amplitude change in a signal. It can be calculated with equation 3.7.

$$AAC = \frac{1}{N} \sum_{i=1}^{N-1} |X_{i+1} - X_i| \quad (3.7)$$

Maximum Fractal Length

The Maximum Fractal Length (MFL) is a way of measuring small changes in muscle activity [26]. This feature calculated by using equation 3.8.

$$MFL = \log_{10} \left(\sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (X_i - X_{i+1})^2} \right) \quad (3.8)$$

Minimum

Minimum (MIN) of each channel contributes as a feature. This feature calculated by using equation 3.9.

$$MIN = \min\{X_1, X_2, \dots, X_N\} \quad (3.9)$$

Maximum

Maximum (MAX) of each channel is another feature for the system. The equation 3.10 was used to calculate the maximum.

$$MAX = \max\{X_1, X_2, \dots, X_N\} \quad (3.10)$$

Standard Deviation

Similar to Min and Max, Standard Deviation (STD) is taken into account as a feature, given in equation 3.11.

$$STD = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 \quad (3.11)$$

In the equation 3.11, μ represents mean of the $\{X_1, X_2, \dots, X_N\}$.

3.2.2 Frequency Domain Feature

The fourier transform is the method to obtain frequency information from a given signal. In this study, the signal collected from subjects are discrete samples. Thus, the Discrete Fourier Transform (DFT) is used. The DFT computed using built in Matlab command “fft”, and then this frequency information divided into predefined bands and the mean of each band is calculated. These coefficients can be found by using equation 3.12, and may be used as the features, where $x[n]$ is a discrete signal with N samples and $F[n]$ is frequency coefficients.

$$F[n] = \sum_{n=1}^{N-1} x[n]e^{(-j2\pi i(kn/N))} \quad (3.12)$$

The frequency domain features are obtained by calculating the absolute value of the FFT coefficients.

3.2.3 Power Spectral Density

The Power Spectral Density (PSD) has smoother distribution comparing with FFT. Therefore, it may reduce the unwanted noise. Welch’s method is used to obtain the PSD. In Figure 3.2, frequency response and the PSD of the signal can be seen.

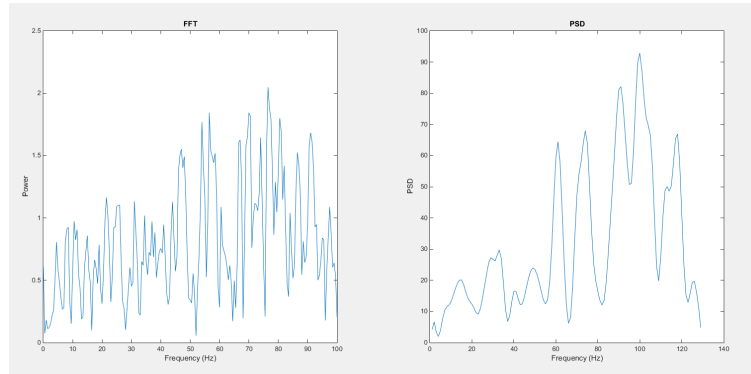


Figure 3.2: Power using absolute value of FFT and PSD using Welch's method

3.2.4 Average Power in Channel

Figure 3.3 shows the human forearm muscle anatomy [27]. As it can be seen, each finger is connected to different muscle groups. Therefore, when a subject performs a gesture, each finger invokes different muscle groups. Thus, the sEMG information will have pattern for each finger movement and this information can be used as features. Figure 3.4 shows a plot of surface EMG signals of the letter 'A' and letter 'B'. As it can be seen that each gesture generates different signals.

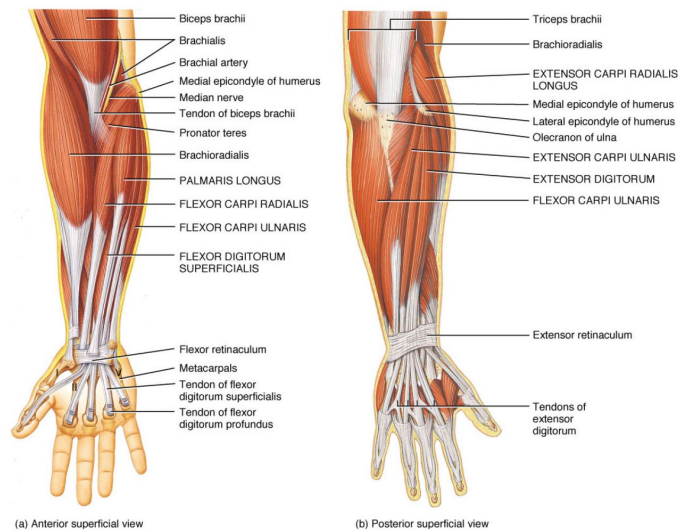


Figure 3.3: The Human forearm muscle anatomy

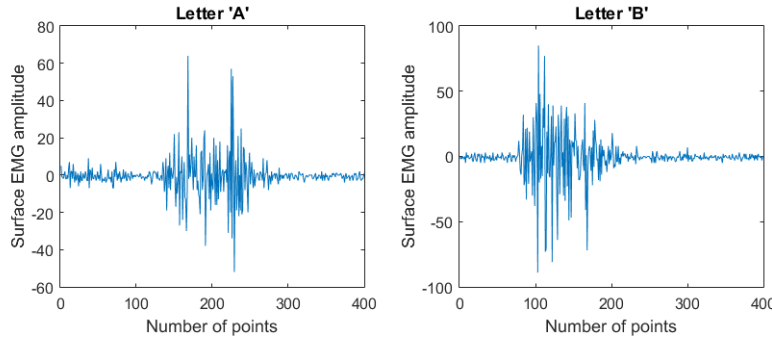


Figure 3.4: A plot of the sEMG of letter 'A' and 'B'

Average power of a channel can be calculated with equation 3.13 where P is the average power.

$$P = \frac{1}{N} \sum_{n=1}^N X[n] \quad (3.13)$$

3.3 Feature Reduction

Principal Component Analysis (PCA) is a feature reduction technique in which a set of possibly related variables are transformed into a set of linearly uncorrelated variables that are called principal components. In other words, PCA is a method used to represent data with fewer number of features. The benefits of PCA include reducing complexity and shrinking the memory size, allowing faster classification. In our scenario, sEMG signals from the forearm were highly correlated due to overlapping of muscles. This is where PCA comes in the picture and reduces the correlation.

Mathematically all PCA does is map all the points in the higher dimensional space to a lower dimensional space by using projection in order to maximize the variance in the samples. The principal components refer to orthogonal eigenvectors. Moreover, the

eigenvalues for each eigenvector correspond to the amount of variance that exists in the data in the direction of the respective eigenvectors [28]. The projection z of a feature vector y onto the column space spanned by eigenvectors W is given by equation 3.14.

$$z = yW \quad (3.14)$$

Our goal is to find a matrix W that is a set of all possible principal components. The optimal solution may be obtained by equation 3.15, where V_L is the L eigenvectors with largest eigenvalues of the covariance matrix.

$$W = V_L \quad (3.15)$$

These principal components are orthonormal vectors which can be calculated using Singular Value Decomposition (SVD). Using SVD, the set of feature vectors Y can be reduced into U , V , and S by equation 3.16, where columns of U are left singular vectors, columns of V are the right singular vectors, and S is a matrix where the main diagonal contains singular values and all other elements are 0.

$$Y = USV^T \quad (3.16)$$

Using SVD, a connection between eigenvectors and singular values can be written as shown in equation 3.17 and 3.18, where $D = S^2$.

$$Y^T Y = VS^T U^T USV^T = V(S^T S)V = VDV^T \quad (3.17)$$

$$(YY^T)V = VD \quad (3.18)$$

The eigenvectors of $Y^T Y$ are equivalent to V , which are the right singular vectors of Y . Moreover, D represents the eigenvalues of $Y^T Y$ and equals the squared singular values. There are some eigenvalues that have little contribution to the variance and may be discarded [29]. Thus, the optimal solution for W can be obtained using $W = V_L$ where V_L is selected from a truncated SVD using a rank L approximation. The value of the L is the minimum number of the eigenvalues whose sum is p percent of the total sum of all eigenvalues. The L is calculated by using equation 3.19.

$$L = \underset{l}{\operatorname{argmax}} \left\{ l \mid \sum_{i=1}^l D(i, i) \leq \operatorname{tr}(D) \cdot p \right\} \quad (3.19)$$

After learning the PCA matrix W , feature vectors are projected into the principal component space to give the reduced feature vector matrix that is used during system model learning and classification.

After feature extraction, 80 time domain features, 96 frequency domain features, 96 PSD features, and eight average power features were obtained over all eight channels. Since only one average power was extracted per channel, PCA was not applied. Table 3.1 shows the number of the features remain after PCA.

Table 3.1: Number of the feature before and after PCA		
Feature Extraction Method	Before PCA	After PCA
Time domain features	80	19
Frequency domain features	96	65
PSD band power features	96	51

Next, we will explore the two classification algorithms were used in this study.

3.4 Classification Methods

There are three learning approaches in machine learning, supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is a method of predicting a function from a labeled dataset. A dataset contains training samples and each sample has features and labeled classes. In other words, supervised learning algorithms analyze the given dataset and produces a predicted function which is called hypothesis. In order to have a good hypothesis, the dataset should contain distinguishable features. A good algorithm is the algorithm that predicts unseen samples correctly. Unsupervised learning is a method that finds the invisible structure in unlabeled dataset such as K Nearest Neighbor [30]. Lastly, reinforcement learning is a type of machine learning that it allows system to automatically determine the ideal action in current circumstances, in order to maximize its performance [29].

In this research, following Machine learning algorithms were used:

- Support Vector Machine
- Ensemble Learning (Bagged Tree)

3.4.1 Support Vector Machine (SVM)

The SVM is one of the supervised learning algorithms in machine learning. The SVM represents samples as points in the space, then divides them into separate groups classes with as large gap as possible as shown in Figure 3.5. The gap between the groups are determined points that are at the borders of the classes. These points at the borders are

called support vectors. The instances that are not support vectors carry no information does not have any affect on solution [29].

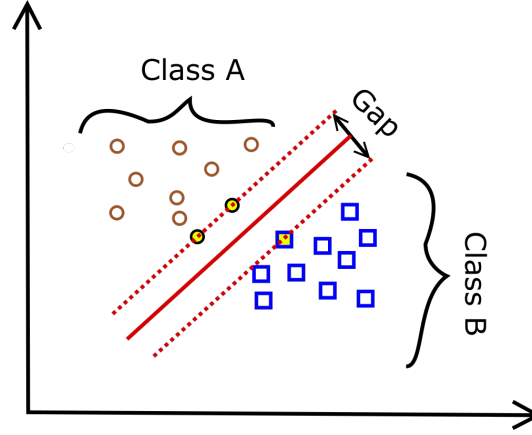


Figure 3.5: Support Vector Machine

Our space is not linearly separable. Thus, a space transformation is needed so that transformed space can be linearly separable. Kernels are used to achieve this in SVM. If a problem is nonlinear, instead of trying to fit a nonlinear model, we can map the problem into a new space by applying a nonlinear transformations using a suitably chosen basis functions and then learn a linear model in this new space [29].

Let assume we have the new dimensions calculated through the basis functions $z = \phi(x)$ where $z_j = \phi_j(x)$, $j = 1, \dots, k$ mapping from the d -dimensional x space to the k -dimensional z space where we write the discriminant as

$$\begin{aligned}
 g(z) &= w^T z \\
 g(x) &= w^T \phi(x) \\
 &= \sum_{j=1}^k w_j \phi_j(x)
 \end{aligned} \tag{3.20}$$

$\phi(x)$ stands for a basis function in equation 3.20.

We also try to use more general case of the soft margin hyperplane here because we have no guarantee that the problem is linearly separable even in this new dimension. The SVM can also be written in equation 3.21 where the $K(x^t, x)$ kernel function.

$$\begin{aligned} g(x) &= w^T \phi(x) = \sum_t \alpha^t r^t \phi(x^t)^T \phi(x) \\ &= \sum_t \alpha^t r^t K(x^t, x) \end{aligned} \quad (3.21)$$

In this study, we have chosen RBF kernels as shown in equation 3.22 where x^t is the center.

$$K(x^t, x) = \exp\left[-\frac{\|x^t - x\|^2}{2s^2}\right] \quad (3.22)$$

To have a good classifier, we need to optimize the SVM parameters, such as sigma, σ . To determine the best sigma value we trained our system with various sigma values and cross-validate the system. After that, we select the value that provides the highest accuracy in both test and cross validation. The green dot in Figure 3.6 shows the best sigma value for given system [31].

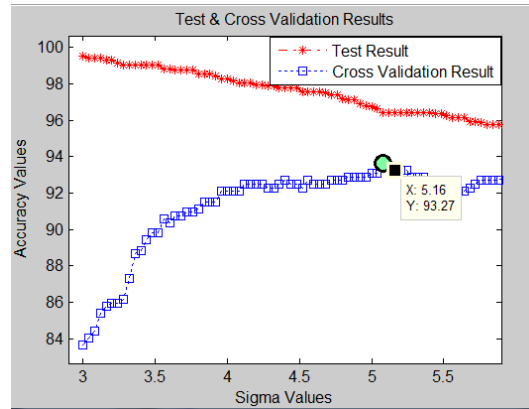


Figure 3.6: Test result vs. Cross Validation result and best sigma value

Next, we will explore our second classification method; Ensemble Learning.

3.4.2 Ensemble Learning (Bagged Tree)

Ensemble learning is a method that creates groups of weak learners and votes their results to obtain better classification as shown in Figure 3.7. One of the ensemble learning algorithms is the Random Forest, proposed by Breiman in 2001 [32]. The Random Forest uses identically distributed decision trees and predict based on the majority of the votes among all decision trees. One disadvantage of the ensemble learning is that it slows down when the number of the trees increase. The ASL dataset consists of 27 classes and these classes overlap. Therefore, ensemble learning would be a good classifier.

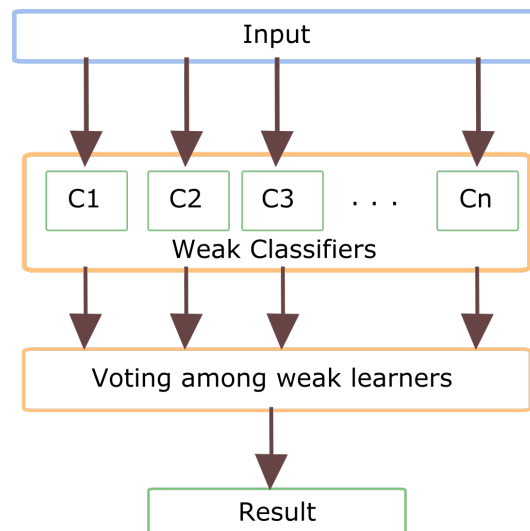


Figure 3.7: Ensemble Learning

3.5 Spell Checker

Mistyping is inevitable while typing a document. To solve this problem, usually text editors use a spell checker. The spell checker either corrects the word or gives users a suggestion of possible corrected options. The proposed method also uses the spell checker to fix the

user's spelling mistakes. Existing spell checkers use different approaches to suggest a word. One of the standard approaches is to calculate the probability of the most used word and to give a suggestion [33]. In this research, existing Microsoft Office Word spell checker is used since the spell checker is not in the scope of this study. However, predicting the word based on confusion matrix of the ASL alphabet classification can be explored to have better spell checker for an ASL system.

Chapter 4

Proposed Method and Experiments

This chapter is organized as follows; data collection hardware, electrode placement, acquiring sEMG data from subjects, and then preprocessing this data set; Next, four different feature extraction methods are applied, time and frequency domain features, power spectral density feature and average power in each channel are extracted; After the feature extraction step, PCA used to find uncorrelated features matrix. The classification section consists of two parts, offline and online system model. In the offline system model, collected dataset were used to learn a model. On the other hand, the online system model does not have to train, it uses pre-learned model. The online system model predicts the ASL alphabet in real-time. Both online and offline system models use SVM and Ensemble Learning. Finally, we present the integration of spell checker and smartphone application to the proposed system.

4.1 Data Collection Hardware

In the experiments, two devices for acquiring surface Electromyography (sEMG) were used. The first one, utilized in the early stage of this research, was Bio Radio 150 by CleveMed. The Bio Radio has up to eight-channel sEMG data reading capability and up to 960 Hz sampling frequency. In addition to the high resolution, it also communicates with a

computer via Bluetooth technology. This feature makes it easier for the subject to move in his/her personal space. Figure 4.1 shows the Bio Radio 150 bio-signal acquisition device [34].

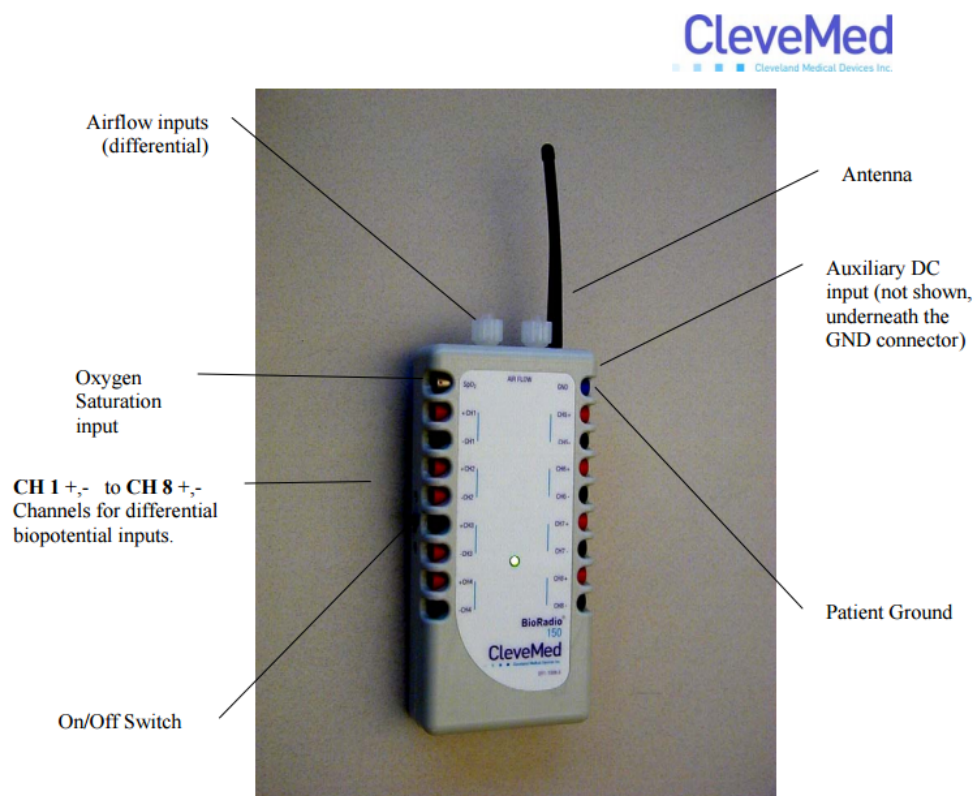


Figure 4.1: The Bio Radio 150 user Unit

Although, the Bio Radio 150 has great sampling rate, there were a few challenges. To get proper sEMG signals, the area under the electrode needs to be cleaned before electrode placement. It also uses sticky electrodes and they are connected to the main unit via cables. To be able to collect eight-channel sEMG signals, 17 different cables were needed. Those wires were causing discomfort for the subjects while performing the sign language gestures. The main disadvantage of the Bio Radio 150 was that it was difficult to achieve the same

placement of the electrodes in different days.

As a solution, the Myo Armband that has eight-channel sEMG electrodes, 9 axis IMU sensors, and built in the battery was used. Myo Armband can be seen in Figure 4.2. The main advantage of this device is that it does not have any cables. It has 200 Hz sampling frequency that is good enough to catch main frequency bands of sEMG signals. In addition, the replacement of the electrodes is no longer a problem because of the special design. It is also very suitable for daily life use.



Figure 4.2: The Myo Armband

4.2 Electrode Placements

The placement of the electrodes has an effect on the quality of the signal collected from the subject's forearm. Therefore, we aimed to choose a place where the device can acquire good signals from muscles. In addition, the subject has to feel comfortable while wearing the device. Thus, the electrodes were placed as shown in Figure 4.3 for Bio Radio 150 and Figure 4.4 for Myo Armband.



Figure 4.3: Bio Radio 150 electrodes placement



Figure 4.4: Myo Armband device placement

4.3 Psychtoolbox 3.0

The Psychtoolbox 3.0 is a computer software that allows researchers to create visual stimulus for their subjects and to collect timed window data of sEMG signals [35]. For this study, a custom data collection system is developed by using the Psychtoolbox 3.0. The data collection system is responsible for two tasks. The first task is reading the sEMG data every 250 ms and buffering the raw data in a sliding window buffer. The second task is displaying the English Alphabets on the screen while buffering the data from the device. This is accomplished by using multi threading in Matlab.

4.4 Data Collection Process

The participants in this study were ten subjects (five male and five female) with ages between 20 and 28 years old. This group included both left-handed and right-handed individuals. The subjects of this research had a diversity of hard of hearing and hearing individuals. A consent form approved by Rochester Institute of Technology Human Subject Research Office (HSRO) was provided for all subjects. The complete application for Human Subject Research filed for this research can be found in Appendix A. Every subject was asked to fill in a consent form. All the questions of the subjects were answered. Then, the instructions that explain simulation steps provided. Although each subject has a different forearm size, Myo Armband was placed similarly on all subjects.

A custom data collection software is implemented by using Psychtoolbox that is mentioned in 4.3. There are three steps in data collection process as it is shown Figure 4.5.

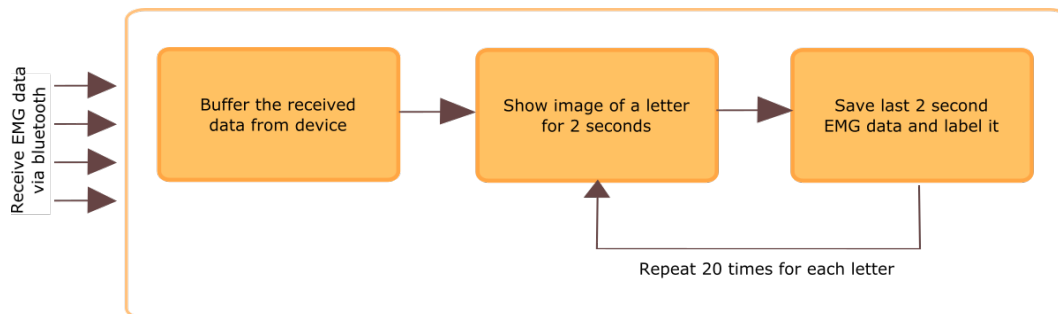


Figure 4.5: Data Collection Diagram

The first step is acquiring data from the device. A custom application was implemented to acquire sEMG data from device and send the buffered data to the system. A screen shoot of the software can be seen in Figure 4.6. In the second step, the letters appear on the screen randomly. An example for letter 'B' can be seen in Figure 4.7. There is

one additional gesture added to the data collection to classify the home position. Subjects were asked to do nothing when they see this extra letter on the screen. The sEMG signal was continuously obtained from the subject and buffered in the system. The visual stimuli shows each American Sign Language letter for two seconds interval and subject was asked to perform the corresponding gesture during this period. Each letter is repeated 20 times in each section. In the third step, the sEMG data were recorded for all alphabets (26 letters and home position) and labeled and saved in files. The data collection consists of two identical sections. The reason for having two sections is to allow the subject to take a long break to rest. They also have a break after every 20 letters if they wish.

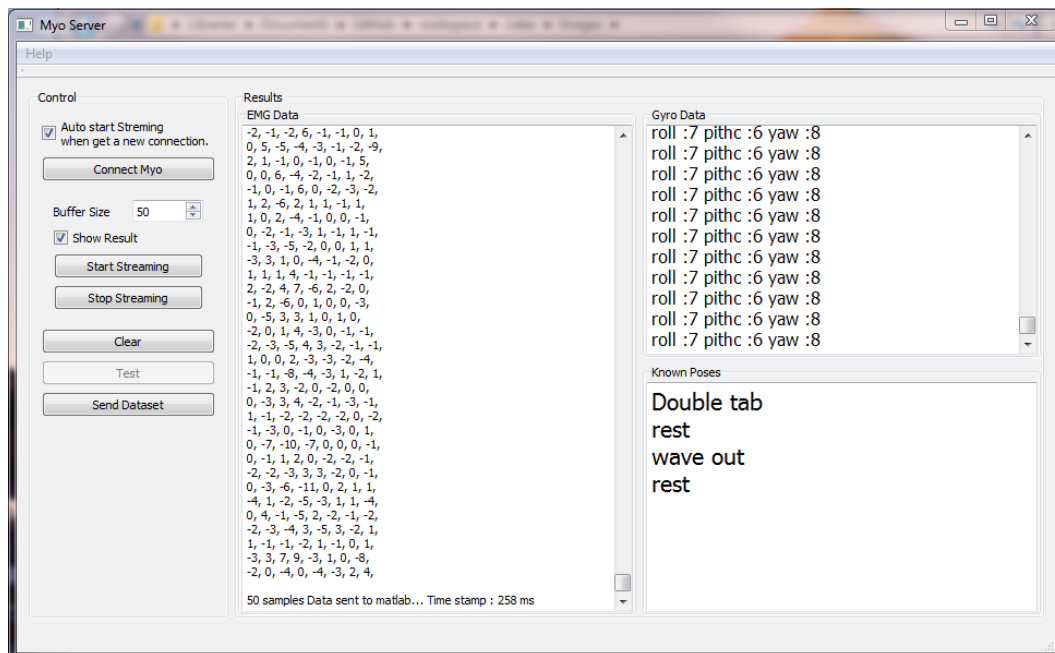


Figure 4.6: A custom software that read sEMG signal from device

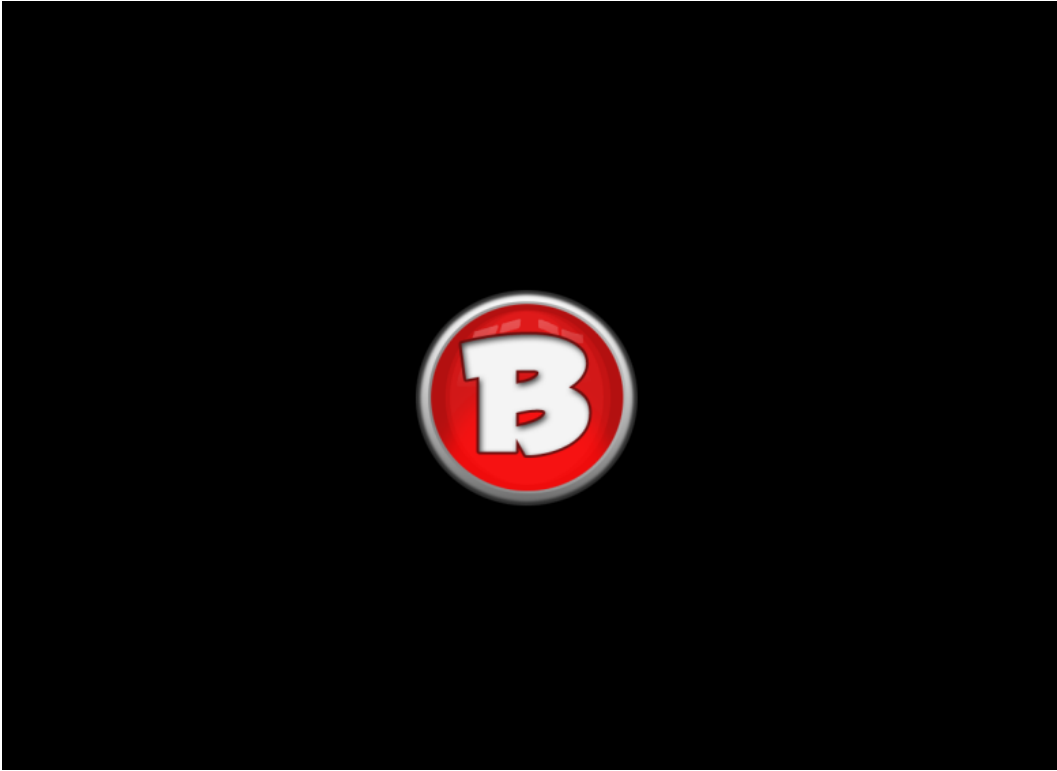


Figure 4.7: A screenshot of the simulation that subject performs the letter 'B'

4.5 Offline System Model

The offline system model consists of four steps as shown in Figure 4.8: preparation, pre-processing, feature extraction, feature reduction, and classification. In the preparation step, because data was collected from the subject on different times it is concatenated to one dataset. The second step is the preprocessing, the methods described in Section 3.1 are used, to remove artifact and unwanted frequency information. Then this preprocessed data were passed to the third step, feature extraction. In this step, time and frequency domain features, power spectral density features and average power in each channel features were applied as we mention in Section 3.2. After feature extraction, we concatenate each channel features to a single row vector where each element of the vector represents a feature,

and the last element presents result. To decrease the computational power and reduce unnecessary features, PCA applied to the feature matrix as the feature reduction step. In the last step, 75% of this dataset was used for training and cross validation using k-fold cross validation technique and rest is used for testing. The k-fold cross validation algorithm divide the data set into a given number of folds. Then it keeps the first part and train the system with rest of the dataset. We select k to be five in this study. After training is done, it cross validates the trained system. This iteration continues for the second part and third so on. After 5-fold cross validation the best system selected.

During offline model learning, Principal Component Analysis (PCA) matrix in the feature extraction and the learned model parameters were saved to be used in the online (Real-time) tests.

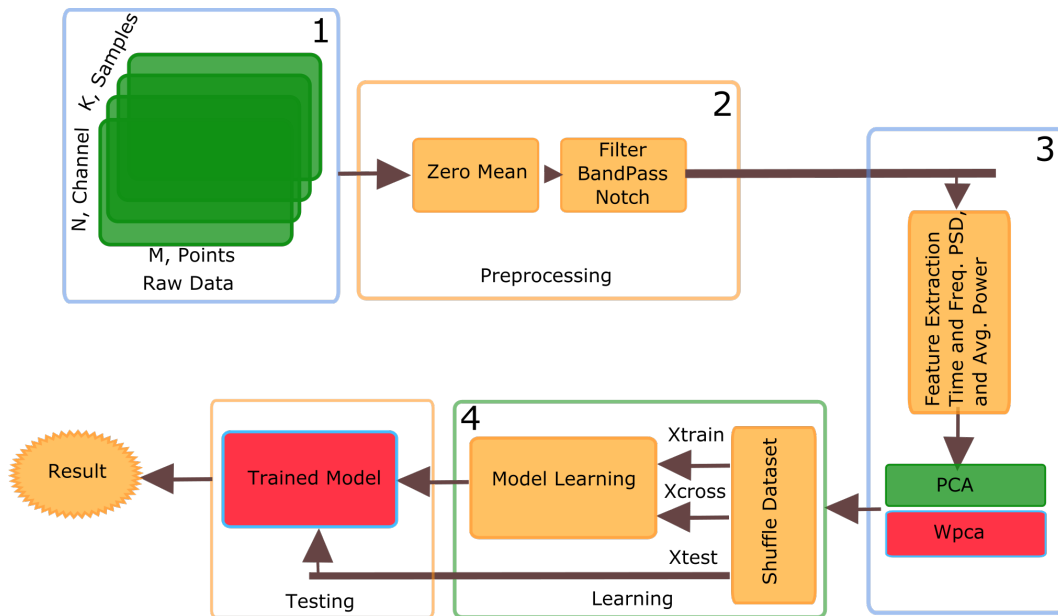


Figure 4.8: Offline American Sign Language System Model

4.6 Online System Model

The online system model does not have the training step because it uses the model parameters of the offline system model as described in Section 4.5.

Whenever new data is classified using the online system model, raw data is preprocessed and feature extraction is performed. As a result of these steps, a row feature vector is obtained. Then these feature vectors go through feature reduction step. The classifier uses these reduced features to predict which class it belongs to using the learned classifier in the offline system model. In Figure 4.9, overall online system model diagram is shown.

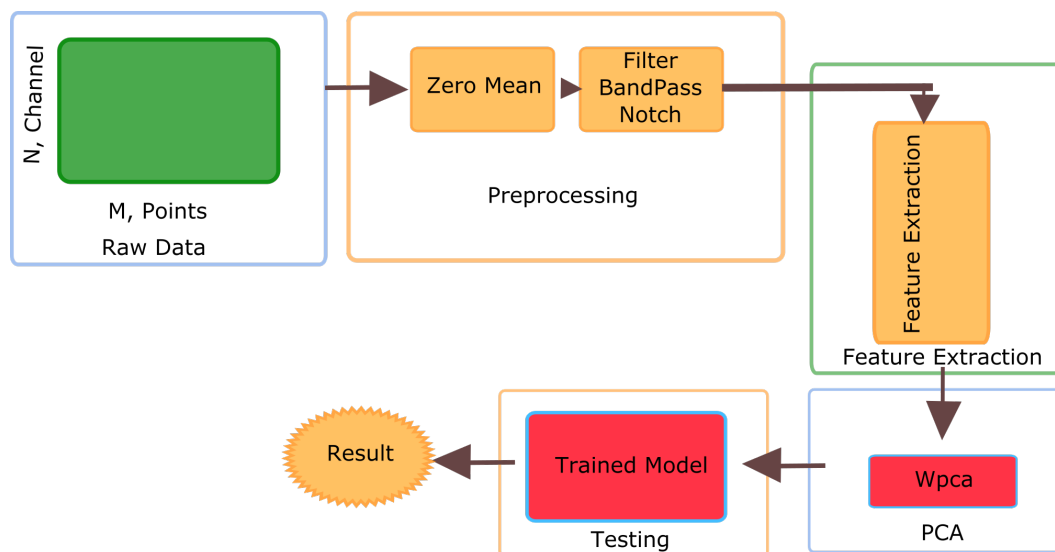


Figure 4.9: Online American Sign Language System Model

4.7 Android Application

Smartphones have been widely used by people for the last 10 years. Thus, most of the population own a smartphone. These devices have powerful CPUs and high memory capacity. Therefore, most of the smartphones can run complex games that require higher

computational resources.

A custom Android application was developed for our proposed ASL system as a proof of concept. The server application is responsible for predicting user's gesture and sending the classification result to the smartphone application. The smartphone application receives the results from the server and shows them on the screen as shown in Figure 4.10.

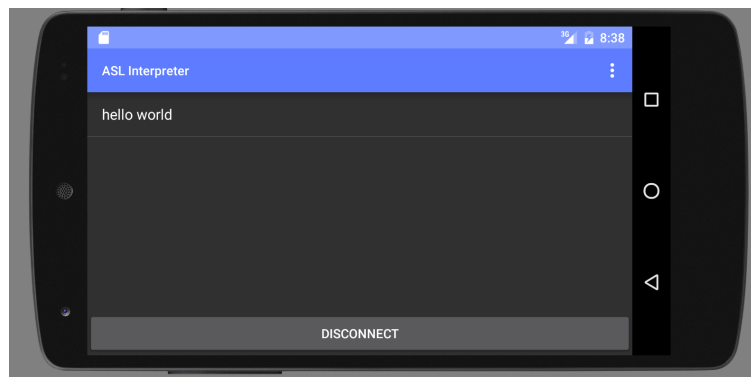


Figure 4.10: Smart Phone Application

In this stage, machine learning algorithm runs on a Matlab application. As we mentioned above, it can run on a smartphone as a standalone application, since smartphones are powerful. However, the standalone application is one of the future work of this thesis.

4.8 Spell Checker

As mentioned in Section 3.5, the proposed method uses the Microsoft Office spell checker engine. The system sends every word to the spell checker to reduce the amount of spelling mistakes. An example can be seen in Figure 4.11. The spell checker fixes the given mistyped word “txst” to “test”.

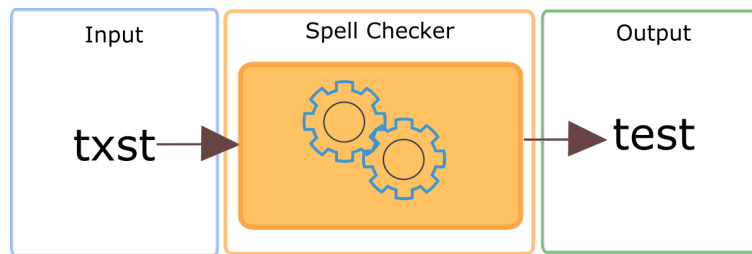


Figure 4.11: Spell Checker Diagram

4.9 Identifiers

The purpose of identifiers is to allow users to have additional control while typing. For example, double tap means user starts typing and system acquires data from the device and predicts the letter and buffers it. Whenever the user performs another double tap gesture, the system takes the buffered letters and sends it to the spell checker. If the user spelled the word correctly, then the system does not change it. However, if there is a typo in the word, then the spell checker replaces it with a correct spelling. This method avoids unnecessary predicting of the sEMG signals if the user does not perform the double tap action. There are other identifiers as well that allow the user to delete a letter and add a space, or a period.

Double Tap : Start/Stop

To be able to start typing with Myo Armband, the double tap has to be performed. First one is interpreted as “start typing” and the second tap as “stop typing”. After every other double tap buffered letters are sent to the spell checker for spell correction. The double tap gesture can be seen in Figure 4.12 [23].



Figure 4.12: Double tap gesture.

Wave in : Delete a letter

The wave in gesture allows the user to remove a letter from the word as shown in Figure 4.13 (a).

Wave out : Space

The wave out gesture adds space as shown in Figure 4.13 (b).

Fist : Add period

The fist gesture adds a period at the end of the word as shown in Figure 4.13 (c).

Wave in, wave out, and fist gestures work only after the second double tap. If a user performs these gestures between first and second tap, the system does not respond to them.

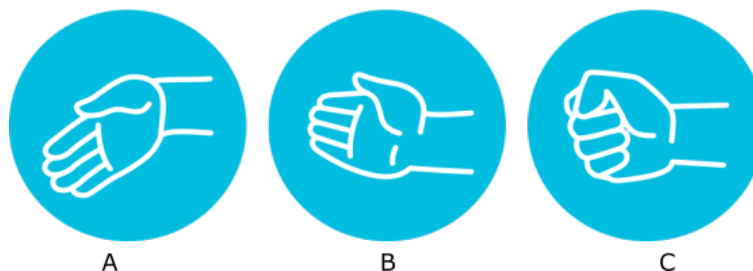


Figure 4.13: (A) Wave in , (B) Wave out and (C) Fist gestures

An example of using identifiers can be seen in Figure 4.14. In the example, the user

decided to type "hello". To write "hello" in the screen, the user has to start with double tap and perform each letter with approximately 2 seconds interval then perform second double tap. If the user does not perform second tap, the system keeps predicting letters and assumes they are part of a single word. The system also ignores the other identifiers if second double tap is not performed.

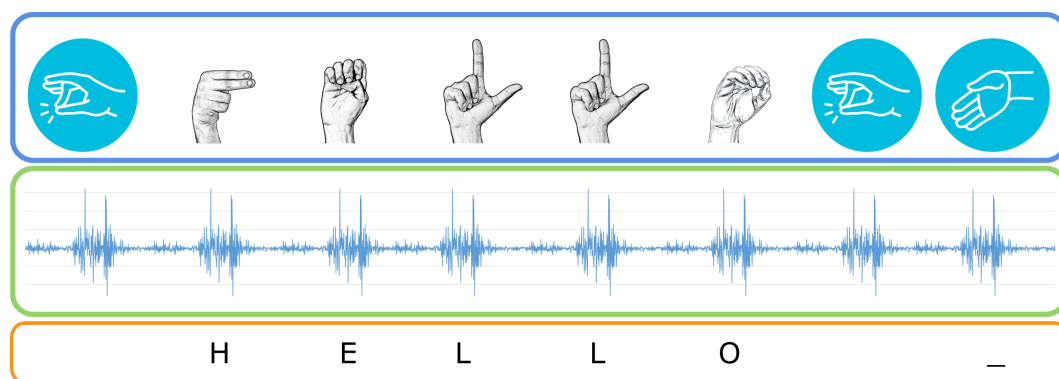


Figure 4.14: How The System Works

Chapter 5

Results and Discussions

This chapter presents results of the experiments for the proposed American Sign Language Recognition System. Experiments are performed with single subject and multiple subjects as mentioned in Chapter 4.

5.1 Single Subject Results

To see the usability of the sEMG signal for the American Sign Language Recognition System, data collected from subjects were applied and validated on the offline ASL system mentioned earlier. The Sign Language Recognition system was explored using various feature extraction methods: time and frequency domain features, PSD band power, and average power in channel and classification methods: SVM and Ensemble Learning.

In the offline ASL system, four feature extraction methods and two classification approaches were implemented. Feature extraction methods mentioned in Section 3.2 are time domain features, frequency information (band power), power spectral density band power and average power in each channel. Then, PCA applied to extracted features to obtain reduced uncorrelated feature matrix. After PCA, the feature matrix is used in each classification algorithm. Analysis of the above mentioned feature extraction methods in

combination with two classification methods is done.

Following steps were taken in all the tests:

1. Dataset were divided in two parts, first part was utilized for training and cross validation set which was 75% of the dataset. Remaining 25% of the dataset was used for testing the system. The test set was never used in the training process.
2. The 75% of the dataset used in 5-fold cross-validation to find the best classification model.
3. When best model found, the system is tested with the test dataset and results are saved.
 - (a) if SVM is used then optimize the sigma by trying different sigmas in range of 1 to 5.
 - (b) if Ensemble is used, set the number of the trees as 200. Figure 5.1 shows that 200 trees is a good number for Ensemble learning.
4. These three steps repeated ten times and the best is taken.

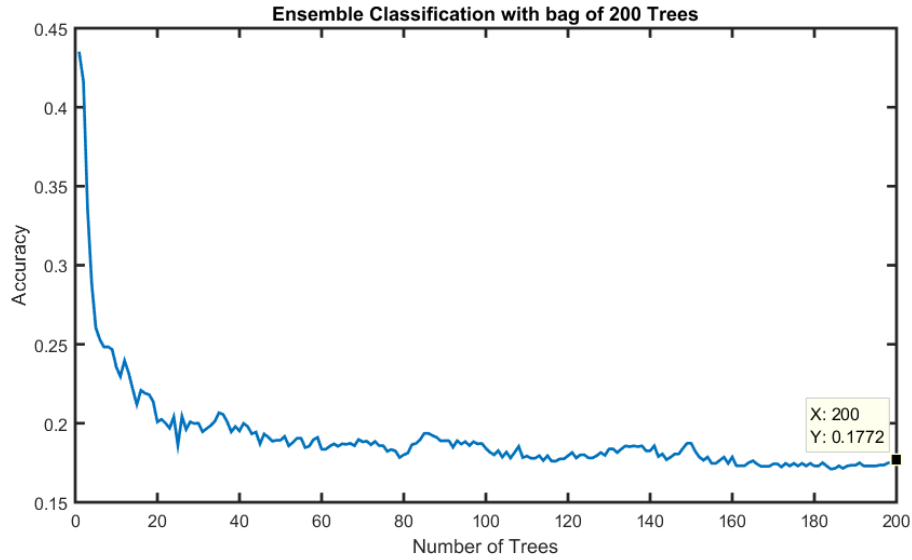


Figure 5.1: The performance of the Ensemble learning using 200 trees

The dataset used for single subject test was collected from a 28 years old subject. The subject repeated 27 letters (one for home position) 20 times in four different days. As a result, each letter repeated 80 times in total. The dataset is processed as mentioned above.

Our first experiment was testing system by using a single subject dataset with SVM classifiers. Figure 5.2 shows the results of 10 iterations of the system. We can see in Figure 5.2 that the SVM was performing well with average power in each channel features. The results of other methods are lower than 50 percent. Although we were assuming that smoothing FFT may provide a better result, FFT's results were better than PSD's results.

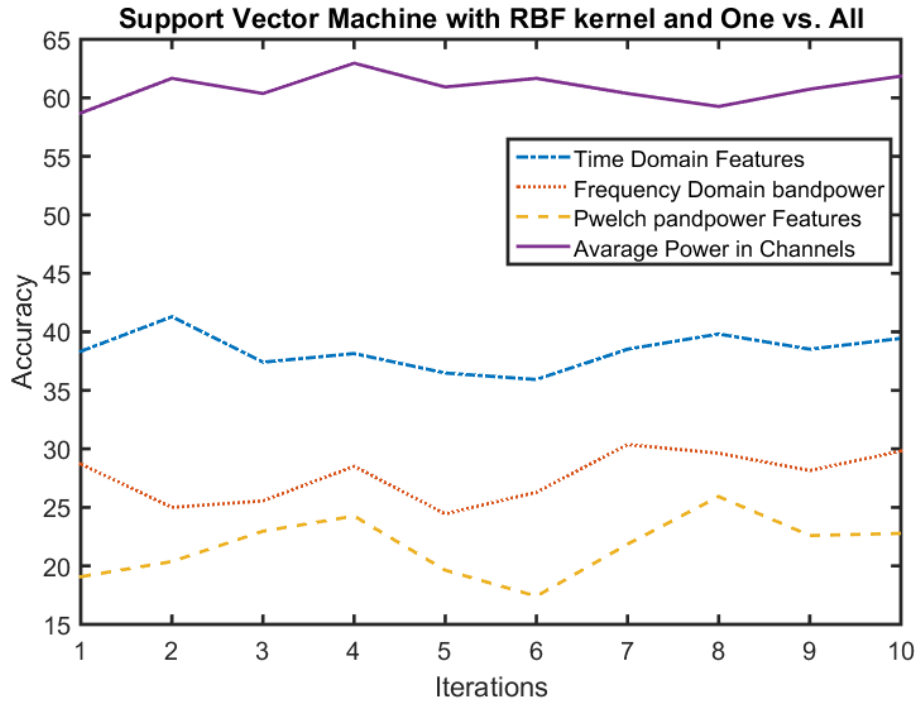


Figure 5.2: Single subject different feature with SVM classifier

Table 5.1 shows the results of the single subject experiment for four feature extraction methods: time domain, frequency domain, PSD band power and average power in each channel. These methods were tested using SVM classifier. The best result obtained from the system was average power in each channel features. Using this method, system reaches 60.85 percent accuracy with a standard deviation of 1.74. The second good feature extraction method was Time domain features that have about 30% accuracy. The worst result obtained from our system was using PSD features.

Table 5.1: SVM, robustness result with different feature extraction methods

Feature Extraction	Accuracy	Standard Deviation
Time domain	38.38	2.11
Frequency band power	27.64	2.66
PSD band power	21.68	1.48
Avarage Power in Channel	60.85	1.74

The second experiment conducted was using the same dataset with a another classifier. In this experiment, we used Bagged Tree classifier which is one of the ensemble learning technique as we mentioned in Section 3.4. Figure 5.3 demonstrates results of the system that tested against same feature extraction methods that used in the SVM. As it can be seen in Figure 5.3, at the first trail, time domain and average power in each channel had a similar result, but if we look at the rest of iterations, we can conclude that average power in each channel was suppressing Time domain features. The PSD was performing worse than Frequency domain features until 7th trail, and then it has started to perform similarly.

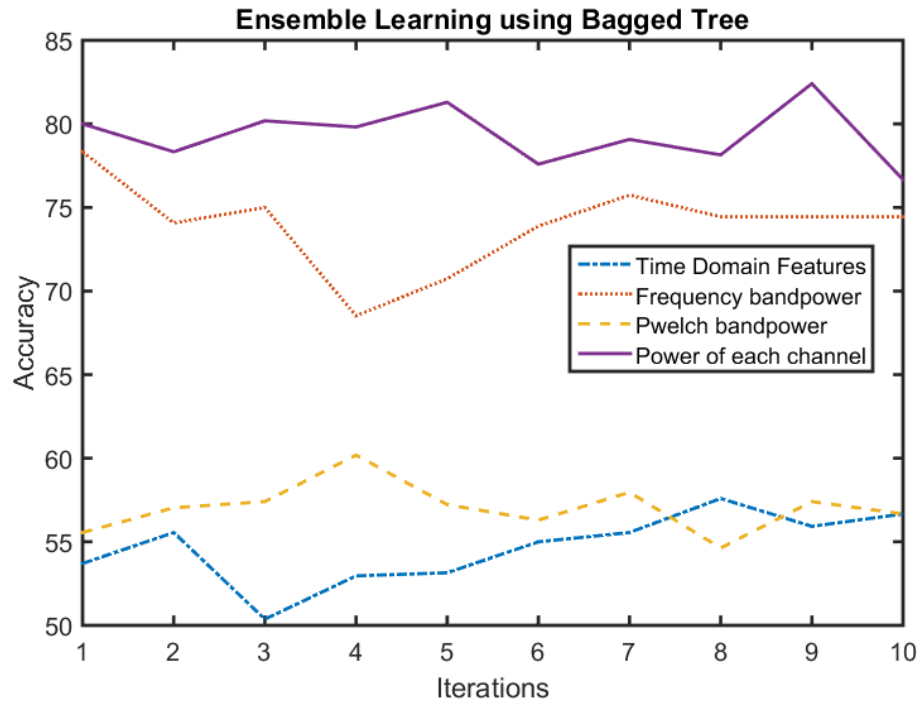


Figure 5.3: Single subject different features with Ensemble classifier

Average power in each channel's features are the best feature for Bagged Trees learner. Table 5.2 shows that it has 79.35% accuracy with a standard deviation of 1.26. The second feature extraction method is the FFT features that reached about 73% accuracy. Once again we experienced that the FFT features are providing better results than PSD features.

Table 5.2: Ensemble Learning robustness result with different feature extraction methods

Feature Extraction Methods	Accuracy	Standard Deviation
Time domain features	54.64	1.57
Frequency domain features	73.96	2.15
PSD band power features	57.03	2.56
Avarage Power in each channel features	79.35	1.26

5.2 Multiple Subject Results

In this experiment, dataset is generated from ten subjects were used. Table 5.3 shows the detailed information about the dataset. Subject Id column contains subject's code. The code S001 stand for subject id and 'H — N' stands for Hard of Hearing/Hearing subjects, 'M — F' stands for Male/Female and 'R — L' stands for Right-handed/Left-handed. Section 1-2 show how many minute a section took. Following column shows how many times each class is repeated. The last column contains the total number of gestures performed by the subjects.

Table 5.3: The sEMG data information of the subjects

Subject ID	Section 1 (Min)	Repetition (I)	Section 2 (Min)	Repetition (II)	Total (Gesture)
S001-NFR	23.92	27x20	22.72	27x20	1080
S002-NMR	26.29	27x20	24.82	27x20	1080
S003-HFR	20.56	27x20	25.80	27x20	1080
S004-HFL	22.52	27x20	22.01	27x20	1080
S005-HFL	24.21	27x20	19.17	27x20	1080
S006-HMR	25.86	27x20	25.82	27x20	1080
S007-HFR	23.31	27x20	25.44	27x20	1080
S008-HMR	25.34	27x20	26.10	27x20	1080
S009-HMR	22.90	27x20	21.02	27x20	1080
S010-HFR	20.29	27x20	20.76	27x20	1080

In the third experiment, we used multi-subject dataset that collected from 10 subjects. Similar to the first experiment, four feature extraction methods mentioned in Section 3.2 and two classifiers referred to in Section 3.4 were used.

Figure 5.4 shows the performance of the SVM classifier for multi-subject dataset experiments. It looks like average power in each channel is still the best feature extraction

method among other.

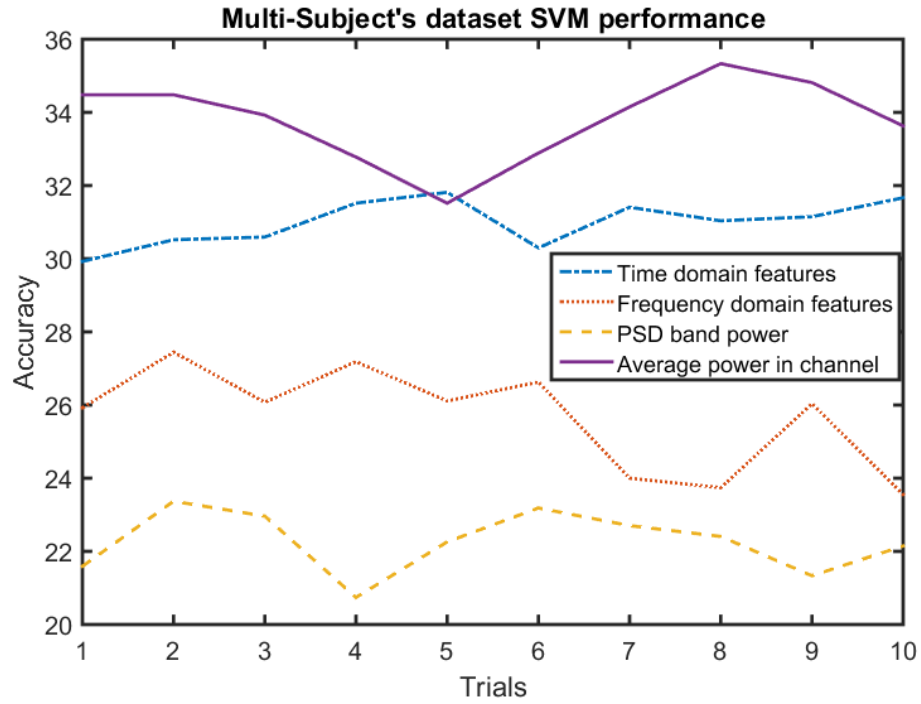


Figure 5.4: Different feature extraction methods with SVM classifier with multi subjects

Table 5.4 shows the results of the SVM classifier for multi-subject dataset. According to results of Table 5.4, we can conclude that SVM is failing with multi-subject dataset. All feature extraction methods performed lower than 34.00% accuracy.

Table 5.4: Multi users, SVM, robustness result with different feature extraction

Feature Extraction	Accuracy	Standard Deviation
Time domain	30.99	0.63
Frequency band power	25.67	1.41
PSD band power	22.27	0.84
Avarage power in channel	33.80	1.13

Figure 5.5 shows the results of 10 trials of the system. In this experiment, Bagged Tree classifier is used instead of an SVM classifier. The number of the subtrees set to 200 trees.

It can be seen that average power in each channel is performing better than others. The second method is FFT features followed by PSD features and Time domain features. Once again we see that FFT features are performing better than PSD features.

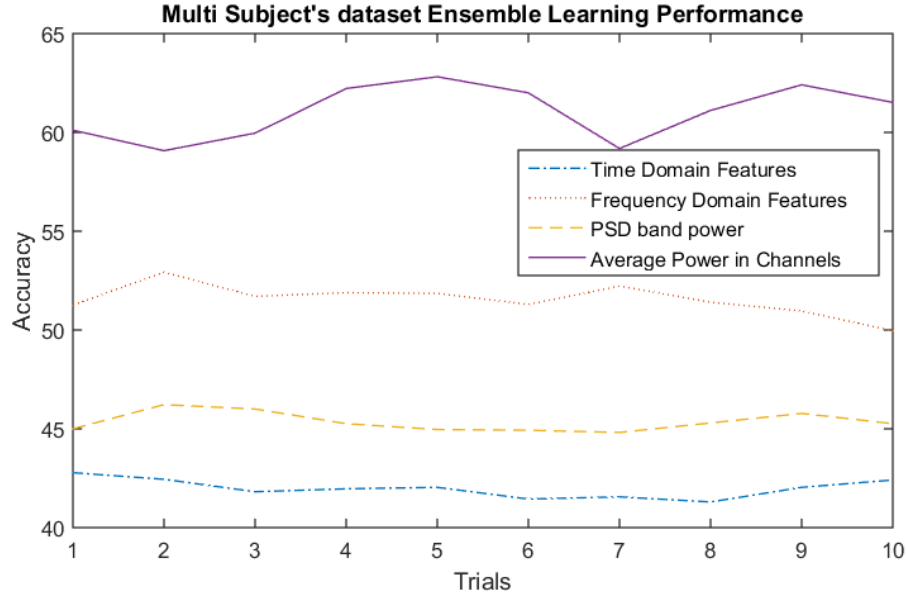


Figure 5.5: Multi subjects different features with Ensemble classifier

In Table 5.5, average power in each channel method performed better than three other feature extraction methods. The accuracy of average power in each channel is 61.04% accuracy with standard deviation of 1.37. The second method is the FFT features with 51.54% accuracy with a standard deviation of 0.78. Finally, the worse performance was obtained from time domain features.

Table 5.5: Multi users, Ensemble Learning, robustness result with different feature extraction methods

Feature Extraction	Accuracy	Standard Deviation
Time domain	41.97	0.47
Frequency band power	51.54	0.78
PSD band power	45.35	0.48
Avarage power in channel	61.04	1.37

The confusion matrix is a way to see individual class performance versus all other classes. Confusion matrix consists of rows and columns that add up to 1. The rows hold the class accuracy across all classes. When we analyze the confusion matrix, we need to look at the diagonal elements. High accuracy at the diagonal elements means it is a good classifier.

Figure 5.6 shows confusion matrix of the single subject dataset trained with Bagged Tree classifier. When we analysis the confusion matrix, it makes more sense why some letter has low accuracy. For example, letter A is 81% times predicted correctly, and it is mixing up with several letters: 7% times with letter I, 4% times with letter K, and 7% times with letter N. When we look at the row G, 75% this the classifier predicts correctly and 15% times confused with letter H. In Figure 5.7, it can easily be seen that letter G and H has similar gesture. Therefore, it makes sense can not differentiate these two letters. As another example, letter R is 15% times mixing up with letter U. These two letters also have similar hand gesture as shown in Figure 5.8. There are other examples can be found in the confusion matrix which has very similar gestures for different letters.

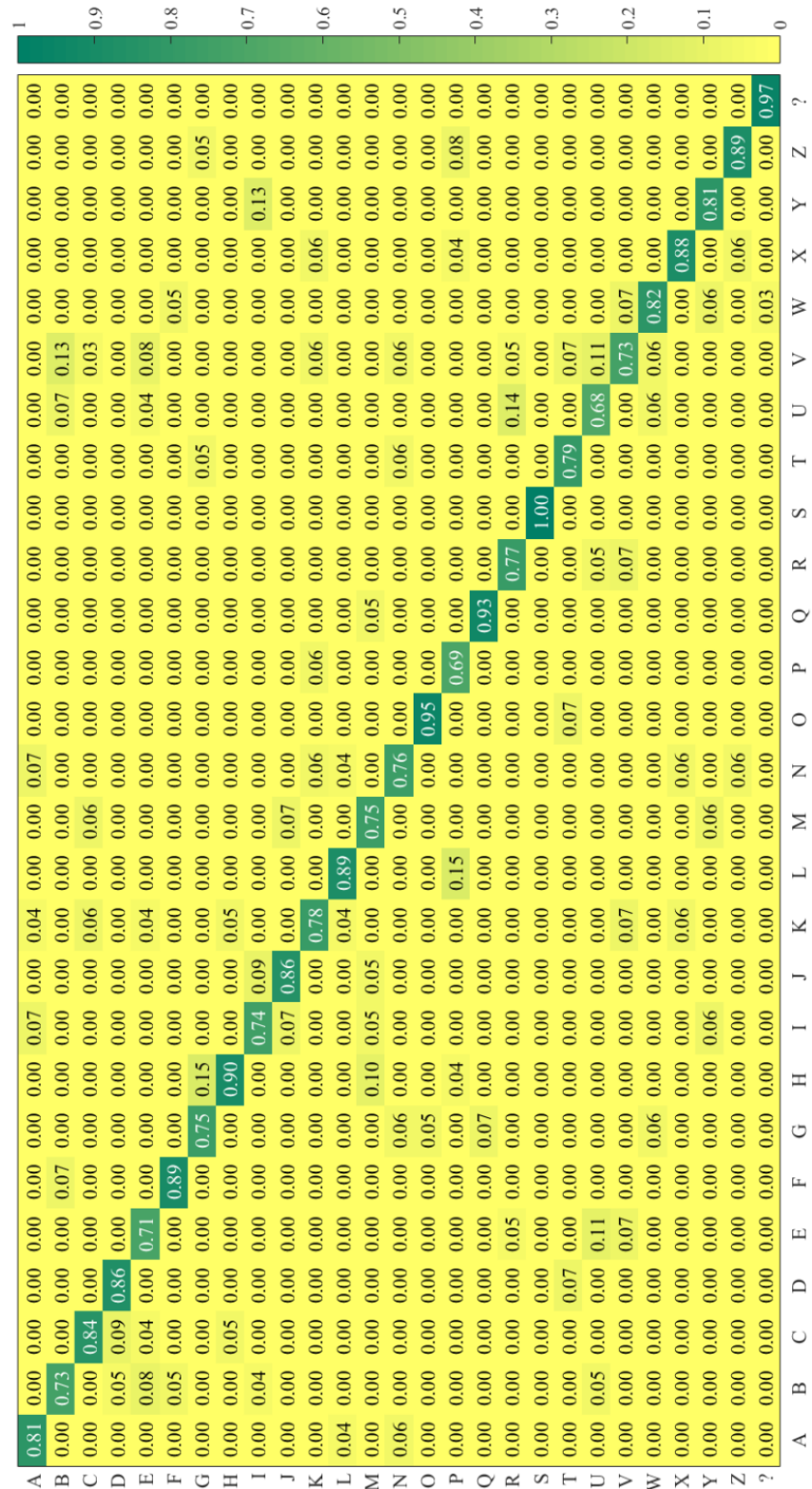


Figure 5.6: The Confusion matrix of the Single Subject Result

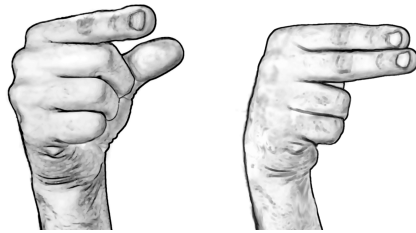


Figure 5.7: Gesture of letter 'G' and 'H'

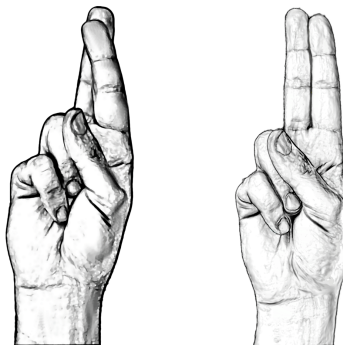


Figure 5.8: Gesture of letter 'R' and 'U'

5.3 Discussions

The results of the experiments in the Chapter 5 show that sEMG signals can be used for Sign Language Recognition System. The proposed system has around 80% accuracy for a single subject. However, the accuracy for multiple subjects not as good as the single subject because of following problems:

1) The electrode placement is not the same for every subject because each person has different forearm size. Although we tried to place the Myo armband on the same location, it is still an approximation. In addition to electrode placement, people have different skin properties affecting the quality of the signal.

2) Each person performs the gestures slightly differently.

3) As there are 27 classes to classify, 26 ASL-English alphabets and a home position, developing a classification system for 27 classes is a challenge. Moreover, the 27 classes are overlapping. Figure 5.9 shows the plot of the randomly selected two features in the dataset.

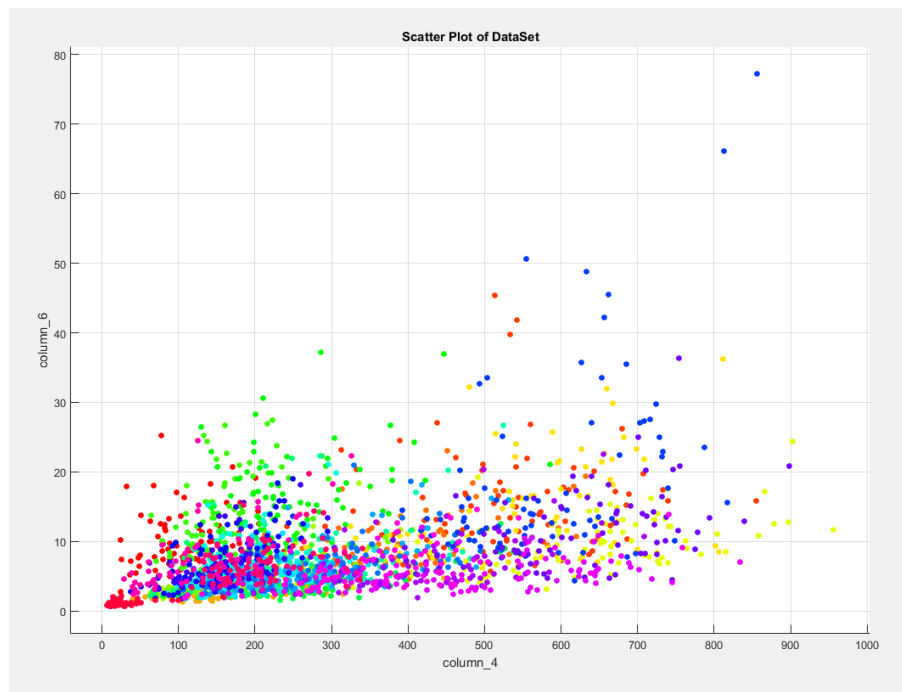


Figure 5.9: 2D (radomly selected two features) representation of the 27 classes dataset

Although, the SVM is a good classifiers, Ensemble Learning provided better accuracies in our experiments. The main reason high performance is Ensemble learning creates many decision trees and votes among them. On the other hand, SVM uses a single classifier for prediction. Since there are 27 classes and they overlap each other, it is really complex problem to solve with an SVM classifier.

Chapter 6

Conclusion

In this study, an American Sign Language Recognition System using Surface EMG signal that allows hard of hearing people to communicate society is presented. To implement the SLR system sEMG signals were collected from multiple subjects. The dataset, modelled and classified, is time series dataset from subjects who know American Sign Language Alphabet. The system blocks have been defined as data collection, preprocessing, feature extraction, feature reduction, and classification. These are the general machine learning procedures. In this work, we have focused more on exploring feature extraction methods and classification algorithms. The feature extraction methods we have explored are time domain features, frequency domain features, power spectral density features, and average power in each channel.

The result of our experiments, which conducted with the single subject dataset and multi-subjects dataset, shows average power in each channel feature is the best feature extraction method among others. We also observed that FFT's band power features are better than PSD's band power features, although we were expecting better performance with PSD features. The problem we are trying to solve is not a trivial one. We are trying to predict correct letter from 27 classes and they are overlapping with each other. Therefore,

ASL recognition system is a complex problem.

In this research, we have explored two classification algorithms; Support Vector Machine and Ensemble Learning classifier. We observed that Bagged Trees classifier was performing better than the SVM classifier in both single subject dataset and multi subjects dataset. The best result we had obtained from Bagged Tree classifier with single subject dataset was about 80% accuracy. Moreover, the best outcome we got with SVM was 60.85% accuracy with single subject dataset. As we used multi-subject dataset, the system reaches 61.04% accuracy with Bagged Tree and 33.80% accuracy with SVM.

In conclusion, the result of the experiments shows that average power in each channel feature is the best feature among other three methods. Moreover, Ensemble Learning (Bagged Tree) classifier performed better than SVM classifier.

Chapter 7

Future Work

In this study, we have used variety of feature extraction methods for classification of ASL Alphabet gestures. There are many other feature extraction techniques out there such as Wavelet, number of the peaks, histogram, etc. Implementing those methods may provide better results.

Implementing a new spell checker based on confusing matrix may increase the accuracy of the overall system since existing spell checker's algorithm is based on keyboard layout and most frequently used words.

In this study, a mobile phone application for American Sign Language has been developed as a proof of concept. The system has three different applications that communicate with each other. The first application acquires raw sEMG data from device and buffers it, then sends it to the Matlab application that handles preprocessing, feature extraction, feature reduction, and classification algorithm. This Matlab application then predicts the letter for the given input sEMG data. Then, the Matlab application sends it to the smartphone application to display. Therefore, implementing standalone Android application that communicates with the device directly and processes the raw data and displays it on the screen would be very handy for users.

In addition to Sign Language Recognition, this system can be used as a new input device for typing in computer software such as Microsoft Office, chatting applications, and email applications.

Bibliography

- [1] W. H. O. (WHO), “Deafness and hearing loss, Report,” *World Health Organization Media Center Fact Sheet*, 2015.
- [2] L. Kin, T. Tian, R. Anuar, Z. Yahya, and A. Yahya, “Sign Language Recognition System using SEMG and Hidden Markov Model,” *Conference on Recent Advances in Mathematical Methods, Intelligent Systems and Materials*, pp. 50–53, 2013.
- [3] T. LEWIN, “Colleges See 16% Increase in Study of Sign Language,” [Online]. <http://www.nytimes.com/2010/12/08/education/08language.html> [Accessed: 9-January- 2016], 2010.
- [4] S. S. Fels and G. E. Hinton, “Glove-talk: a neural network interface between a data-glove and a speech synthesizer,” *IEEE Transactions on Neural Networks*, vol. 4, no. 1, pp. 2–8, 1993.
- [5] R.-H. Liang and M. Ouhyoung, “A Real-time Continuous Alphabetic Sign Language to Speech Conversion VR System,” *Computer Graphics Forum*, vol. 14, no. 3, pp. 67–76, 1995.
- [6] W. J. W. Jiangqin, G. W. G. Wen, S. Y. S. Yibo, L. W. L. Wei, and P. B. P. Bo, “A simple sign language recognition system based on data glove,” *ICSP '98. 1998 Fourth International Conference on Signal Processing (Cat. No.98TH8344)*, vol. 2, pp. 1257–1260, 1998.
- [7] S. A. K. Mehdi Y. N., “Sign language recognition using sensor gloves,” *Proceedings of the 9th International Conference on Neural Information Processing*, vol. 5, pp. 2204–2206, 2002.
- [8] M. M. Sole and M. S. Tsoeu, “Sign Language Recognition using the Extreme Learning Machine,” *IEEE Africon 2011 - The Falls Resort and Conference Centre, Livingstone, Zambia*, pp. 13–15, 2011.
- [9] H. E. Hayek, J. Nacouzi, and P. O. B. Z. Mosbeh, “Sign to Letter Translator System using a Hand Glove,” *The Third International Conference on e-Technologies and Networks for Development (ICeND2014)*, pp. 146–150, 2014.

- [10] G. W. Vicars, "American Sign Language Alphabet," [online]. <http://www.lifeprint.com/asl101/fingerspelling/abc-gifs/index.htm>. [Accessed: 9-January- 2016], 2015.
- [11] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998.
- [12] K. O. Rodriguez and G. C. Chavez, "Finger spelling recognition from RGB-D information using kernel descriptor," *Proceedings of Brazilian Symposium of Computer Graphic and Image Processing*, pp. 1–7, 2013.
- [13] C. Keskin, F. Kırç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," *IEEE International Conference on Computer Vision Workshops*, pp. 1228–1234, 2011.
- [14] C.-H. Chuan, E. Regina, and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," *2014 13th International Conference on Machine Learning and Applications*, pp. 541–544, 2014.
- [15] K. S. Raut, "Recognition of American Sign Language Using LBG Vector Quantization," *2014 International Conference on Computer Communication and Informatics (ICCCI -2014)*, pp. 2–6, 2014.
- [16] M. A. Soltanmoradi, V. Azimirad, and M. Hajibabazadeh, "Detecting finger movement through classification of electromyography signals for use in control of robots," *Proceeding of the 2nd RSI/ISM International Conference on Robotics and Mechatronics*, pp. 791–794, 2014.
- [17] J.-U. C. J.-U. Chu, I. M. I. Moon, Y.-J. L. Y.-J. Lee, S.-K. K. S.-K. Kim, and M.-S. M. M.-S. Mun, "A Supervised Feature-Projection-Based Real-Time EMG Pattern Recognition for Multifunction Myoelectric Hand Control," *IEEE/ASME Transactions on Mechatronics*, vol. 12, no. 3, pp. 282–290, 2007.
- [18] T. S. Saponas, D. S. Tan, D. Morris, J. Turner, and J. A. Landay, "Making muscle-computer interfaces more practical," *Proceedings of the 28th international conference on Human factors in computing systems*, pp. 851–854, 2010.
- [19] T. S. Saponas, D. S. Tan, D. Morris, and R. Balakrishnan, "Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces," *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems*, pp. 515–524, 2008.

- [20] M. S. Erkilinc and F. Sahin, "Camera control with EMG signals using Principal Component Analysis and support vector machines," *2011 IEEE International Systems Conference (SysCon)*, pp. 417–421, 2011.
- [21] U. Sahin and F. Sahin, "Pattern recognition with surface EMG signal based wavelet transformation," *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 295–300, 2012.
- [22] V. E. Kosmidou, L. J. Hadjileontiadis, and S. M. Panas, "Evaluation of surface EMG features for the recognition of American Sign Language gestures," *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, vol. 2, no. 4, pp. 6197–6200, 2006.
- [23] Thalmic Inc., "Myo Armband," [Online:]. Available: <https://www.myo.com/>, 2015.
- [24] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition," *Journal of Computing*, vol. 1, no. 1, pp. 71–80, 2009.
- [25] D. Tkach, H. Huang, and T. a. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition.," *Journal of neuroengineering and rehabilitation*, vol. 7, p. 21, 2010.
- [26] T. Portfolio, *Fractal features of Surface Electromyogram : A new measure for low level muscle activation*. PhD thesis, 2008.
- [27] B. Tortora J., Gerard, Derrickson, *Principle of Anatomy and Physiology*. Wiley, 14th ed., 2013.
- [28] R. M. Bowen, *Online Novelty Detection System: One-Class Classification of Systemic Operation*. PhD thesis, 2015.
- [29] Ethem Alpaydin, *Introduction to Machine Learning*. The MIT Press, 3rd ed., 2014.
- [30] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbour Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.
- [31] C. Savur and F. Sahin, "American Sign Language Reconition System by Using Surface EMG Signal," in *International Machine Learning and Application Conference ICMLA*, 2015.
- [32] L. Breiman, "Random Forests," *Journal of Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

- [33] P. Norvig, “How to Write a Spelling Corrector,” *[Online]*. Available: <http://norvig.com/spell-correct.html>. [Accessed: 10- January- 2016].
- [34] P. Station and L. Controller, *Bio Radio 150 User’s Guide*. 2011.
- [35] D. Brainard, “The psychophysics toolbox,” *Spatial vision*, pp. 433–436, 1997.

APPENDIX A

Multi Agent Bio-Robotics Laboratory (MABL) **EEG, EMG and EOG Collection**

mablbr@rit.edu

MABL Consent Form Template

Project Title: *American Sign Language Recognition System by Using Surface EMG Signals*

Lab Director: *Dr. Ferat Sahin*
Electrical and Microelectronic Engineering,
Rochester Institute of Technology,
Contact: feseee@rit.edu

Principal Investigator: *Celal SAVUR*
Electrical Engineering
Rochester Institute of Technology
cs1323@rit.edu

Faculty Advisor: *Dr. Ferat Sahin*
Electrical Engineering
Rochester Institute of Technology
feseee@rit.edu

What the study is about

You are being asked to participate in a research study conducted by **Celal Savur** of Rochester Institute of Technology.

The purpose of this research is to use skin electrodes on your muscles to measure electrical signals resulting from muscle movement. This special technique is called Electromyography (EMG). Your participation will help us gain knowledge about American Sign Language.

Recording procedures are NON INVAISVE .i.e. Skin Contact based and no breach of the skin.

Participant ID:

Investigator Initials:

i

What we will ask you to do I will ask you to come to the Multi Agent Bio-robotics Laboratory (MABL) located in Room No. 3271 in Building 9, Rochester Institute of Technology, Rochester, NY -14623 .

[EMG Module] There you will be attached adhesive silver-plated snap electrodes or gold-plate skin electrodes on the concerned muscles in a non-invasive manner. You will be asked to perform showing American Sign Language Letters on the computer screen. Your participation is expected to take a total of 90 minutes, and involve 1 visit.

Risks and Discomforts

The rare and minor physical risks associated with participating in this research are:

- a) Possibility of discomfort or muscle strain in case of body movement as part of the experimental setup.
- b) Possibility for you to receive unperceivable electrical signals from the equipment. The likelihood of you experiencing electric discharge from the Myo Armband equipment is extremely small, as is the potential amount of electricity that you could receive.
- c) Finally, despite careful precautions, there is a risk that your personally identifying information, including measurements we make and the log of our participation in this study, could become available to an unauthorized third party.

Benefits

There are no direct benefits associated with participating in research that involve measurements of EMG.

Information from this study may help us have better classification for American Sign Language. There are many deaf people in the society they cannot use their hand to type. With this project their gesture will be translated to text.

Compensation There will be no compensation this time.

[Video Recording] A video will be recorded of your hand movements, in order to determine the corresponding EMG signal, and will not be associated with your personal information. The records of your movements will be kept for 5 year(s), after which point, considering research-related factors, it may be destroyed. This movement information will be used only as a reference or veracity check of the EMG signals recorded and no identifiable information will be published.

Confidentiality Your participation in this study will remain confidential, and your identifying information will not be stored with your data or images. Servers and computers where the data and images are stored are password protected. Any paper surveys will be kept in locked rooms. Your images and data will be assigned a code number, which will be used in place of your name to allow linkage of data if follow up is

necessary. The list connecting your name with this number will be kept in a locked room. We will review our data storage requirements periodically to determine if data need to be discarded. Only people authorized by the Lab Director and Principle Investigator will be granted access to the data. We may also need to collect some identifying information for administrative purposes (i.e., for study compensation, and/or an unexpected finding report), but this will not be linked to the research records. In the unlikely event of an emergency, we will also need to provide your information to medical and/or emergency assistance personnel. The data will be used for research and educational purposes, such as teaching, publications, and/or presentations and may be viewed by students, other trainees, and professional colleagues. In any sort of report we make public, we will not include any information that will make it reasonably possible to identify you.

Right to Withdraw: Taking part in this study is completely voluntary. If you have no objection to a question, then please answer every question. It is better to answer than to leave it blank (even if you are unsure). If you decide not to take part or not to complete the study, there is no penalty or loss of benefits, and it will not affect your current or future relationship with Rochester Institute of Technology. If you decide to take part, you are free to withdraw at any time.

If you are injured by this research ...

Emergency medical care is not available on-site. In case of such an emergency the RIT public safety will be immediately informed and the necessary help needed will be provided immediately.

But Cost for such care will be billed in the ordinary manner to you or your insurance company. No reimbursement, compensation, or free medical care is offered by Rochester Institute of Technology. If you think that you have suffered a research-related injury, contact **Celal Savur** right away at (***)-***-****.

Withdrawal by investigator, physician, or sponsor: The investigators, physicians or sponsors may stop the study or take you out of the study at any time should they judge that it is in your best interest to do so, if you experience a study-related injury, or if you do not comply with the study plan. They may remove you from the study for various other administrative and medical reasons. They can do this without your consent.

If you have questions Please ask any questions you have now. If you have questions later, you may contact the principal investigator **Celal SAVUR** at mablbr@rit.edu or at (***)-***-****. If you have any questions or concerns regarding your rights as a subject in this study, you can contact the Human Subjects Research Office (HSRO) and the Institutional Review Board (IRB) at RIT (access to their website at http://www.rit.edu/research/hsro/about_contact.php).

We would like you to initial the boxes below to show that you consent to various aspects of the EMG **recording** procedure. Then sign the form at the bottom.

Participant ID:

Investigator Initials:

iii

Please note that signing this consent form indicates your consent for EMG to be recorded.

(To be filled by the subject)

Subject Enrollment Criteria: *(Subject initials in the)*

a) *I have Normal Vision or Normal to Corrected Vision*

b) *I confirm that I have read the following medical questions.
If the answer to any question is "yes", I have discussed with the investigator.*

Answers to the Medical Questions described below:

- *Are you currently taking, or have you recently taken, any prescription or over-the-counter medicine? _____*
- *Have you ever suffered from epilepsy? _____*
- *Have you had surgery in which metal items have been placed in your head? _____*
- *Do you have a heart pacemaker fitted? _____*
- *Do you use any other medical device? _____*
- *Have you been feeling unwell over the last few days? _____*
- *Do you suffer from any chronic skin condition (e.g. dermatitis, eczema, psoriasis)? _____*
- *Have you consumed any alcohol or recreational drug over the last 24 hours? _____*
- *Do you suffer from any condition impairing blood clotting (e.g. hemophilia) or are you taking any medication that could affect blood clotting? _____*
- *Do you currently have any cuts or abrasions on your head? _____*

c) (For Physically Disabled) *I consent that I have informed the researcher of my physical disability, and that it will be recorded.*

You can choose not to answer any of these questions. If you feel uncomfortable in discussing such information, you should not participate. You do not need to tell us why you have chosen not to participate.

The answers to these questions are for the investigator to judge and make an informed decision in your best interest to participate in the research/study.

Participant ID:

Investigator Initials:

iv

Thank you for your interest in taking part in EMG research.

You will be given a copy of this form to keep for your records.

Participant's declaration

I give my informed consent to participate in the EMG recording session. I am aware that my participation is voluntary and that I may withdraw at any time without giving a reason. I am aware that all information given by me or data recorded from me will be handled confidentially.

Participant Name (printed) _____

Participant Signature _____

Date _____

Investigator's (Researcher's) declaration

I believe the participant has been fully informed about the EMG signal recording procedure to the level necessary for the giving of informed consent. I have discussed all relevant aspects of the procedure with the participant, and answered all questions to their satisfaction. I have observed the participant initials in the Enrolment Criteria Section of this MABL – Consent Form, and have indicated that the Participant *[participant id]* can participate in the study.

Investigator/Researcher Name (print) _____

Investigator/Researcher Signature _____

Date _____

This consent form will be kept by the researcher for at least ____ year(s) beyond the end of the study and was approved by the IRB on _____ (IRB # _____).

I.

MABL Lab

HSRO/IRB Initial Request for Approval:

Risks and Benefits

1. From the list below, please select ALL of the potential risks that are involved in your study.

☐ Risk of injury or bodily harm

☐ Other risks (please specify)

- i. "Risk of mild skin irritation; risk of breached confidentiality; risk of discomfort; risk of eyestrain; risk of exposure to imperceptible electric discharge from EMG equipment."

2. Describe the nature and degree of the risks or harms selected above. All of the risks harms must be disclosed in the consent form.

☐ 'Participants involved in an EEG, EMG or EOG study may experience skin irritation from the use of electrodes, a mild electrolyte gel, and/or adhesive discs. Participants may also experience discomfort as a result of sitting still and minimizing self-adjustments, during experiments lasting up to 1 hour. They may also experience discomfort from having the aforementioned electrolyte gel in their hair; however, it is water-soluble and easily rinsed out of one's hair. The potential for eye-strain due to the constant focusing and tracking on a computer screen or otherwise is extremely unlikely as long as the distance of the screen from the participant is greater than 80 cm, the short duration of the experiment, and the natural protective abilities of the eye. The potential to experience perceptible electric discharge from the EEG equipment is extremely small, as there is only one electrode (out of 20) that carries a return current to the participant, and that electrical current is inherently limited and is lower than the thresholds of perceptible electric discharge of 1 milliamp. Finally, despite careful precautions, the on-site storage of personally identifying information, including names, and recorded bio-signal data, pose a small risk that participant confidentiality could be breached.'

3. Explain what steps will be taken to minimize risks or harms and to protect subjects' welfare. If the study will include protected populations, please identify each group and provide an explanatory paragraph for each group.

☐ 'Participants will be notified verbally, and in the informed consent document, that there are small risks of skin irritation, discomfort, and breaches of confidentiality. Participants will be asked to self-determine their susceptibility to skin irritation and discomfort. The data is to be collected for a maximum span of ____minutes. In order

to avoid discomfort from the electrode cap, strain on eyes for concentrating on the computer screen for a long time, or fatigue from muscle movements, the data collection procedure will be divided in 3 sessions of 30 minutes (less than 30 minutes), in between which the participant could take a break to rest his/her eyes and continue when he/she is ready. At any time during the data collection session, if the subject is uncomfortable or is feeling any kind of discomfort, the session will be immediately stopped. In order to preserve the confidentiality of the subject, the data recorded will not contain any identifiable information or markers. Any identifiable information taken from the subject will be stored on a secure lab server, which is password protected, has firewalls and is connected on to the secure RIT network. Any paper information will be stored in locked cabinets inside approved-card-access-only rooms, and will only be accessible to the Principle Investigator. This information will not be disclosed and will be available only to the investigators of this study.'

References:

IRB Application Templates: **Human Development EEG and Psychophysiology Laboratory, Cornell University.**

<http://www.human.cornell.edu/hd/eeglab/>